

Sowada, Moritz G.

Expertenurteile – Achillesferse oder Trumpf der Schulinspektion?

Pietsch, Marcus [Hrsg.]; Scholand, Barbara [Hrsg.]; Schulte, Klaudia [Hrsg.]: *Schulinspektion in Hamburg. Der erste Zyklus 2007 - 2013: Grundlagen, Befunde und Perspektiven.* Münster [u.a.] : Waxmann 2015, S. 137-156. - (HANSE - Hamburger Schriften zur Qualität im Bildungswesen; 15)



Quellenangabe/ Reference:

Sowada, Moritz G.: Expertenurteile – Achillesferse oder Trumpf der Schulinspektion? - In: Pietsch, Marcus [Hrsg.]; Scholand, Barbara [Hrsg.]; Schulte, Klaudia [Hrsg.]: *Schulinspektion in Hamburg. Der erste Zyklus 2007 - 2013: Grundlagen, Befunde und Perspektiven.* Münster [u.a.] : Waxmann 2015, S. 137-156 - URN: urn:nbn:de:0111-pedocs-115321 - DOI: 10.25656/01:11532

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-115321>

<https://doi.org/10.25656/01:11532>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Marcus Pietsch, Barbara Scholand,
Klaudia Schulte (Hrsg.)

Schulinspektion in Hamburg

Der erste Zyklus 2007–2013:
Grundlagen, Befunde,
Perspektiven



HANSE
Hamburger Schriften zur Qualität
im Bildungswesen

herausgegeben von
Detlef Fickermann und Knut Schwippert

Band 15



Waxmann 2015
Münster • New York

Marcus Pietsch, Barbara Scholand,
Klaudia Schulte (Hrsg.)

Schulinspektion in Hamburg

Der erste Zyklus 2007–2013:
Grundlagen, Befunde und Perspektiven



Das gedruckte Buch finden Sie [hier](#).



Waxmann 2015
Münster • New York

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet über <http://dnb.d-nb.de> abrufbar.

HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 15

ISSN 1864-2225

ISBN 978-3-8309-3278-9

© Waxmann Verlag GmbH, 2015
Steinfurter Straße 555, 48159 Münster

www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Pleßmann Design, Ascheberg
Umschlagfoto: © Robert Kneschke – Fotolia.de
Lektorat und Satz: Judith Zimmer, Hamburg
Druck: Mediaprint, Paderborn

Gedruckt auf alterungsbeständigem Papier,
säurefrei gemäß ISO 9706

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlages
in irgendeiner Form reproduziert oder unter Verwendung elektronischer
Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Inhalt

Vorwort.	9
---------------	---

Teil 1: Die Grundlagen der Schulinspektion

Norbert Maritzen

Schulinspektion – Aspekte der Transformation von Governance-Strukturen . . .	13
------------------------------------------------------------------------------	----

Norbert Maritzen & Herbert Altrichter

Schulinspektion in Hamburg – ein Gespräch über Startbedingungen und Entwicklungen	37
------------------------------------------------------------------------------------------------	----

Martina Diedrich

Aufbau und Rolle der Schulinspektion Hamburg	57
--------------------------------------------------------	----

Thomas Brüsemeister

Erfahrungsbereiche der Hamburger Schulinspektion – Governance-analytische Programmtheorie	77
--------------------------------------------------------------------------------------------------------	----

Susanne Ulrich

Das Qualitätsleitbild der Schulinspektion Hamburg	101
-------------------------------------------------------------	-----

Teil 2: Die Arbeit der Schulinspektion

Marcus Pietsch, Ann-Katrin van den Ham & Olaf Köller

Wirkungen von Schulinspektion: Ein Rahmen zur theoriegeleiteten Analyse von Schulinspektionseffekten	117
-------------------------------------------------------------------------------------------------------------------	-----

Moritz G. Sowada

Expertenurteile – Achillesferse oder Trumpf der Schulinspektion?	137
----------------------------------------------------------------------------	-----

Knut Schwippert

Daten für die Schulentwicklung – auf die Perspektive kommt es an	157
----------------------------------------------------------------------------	-----

<i>Barbara Scholand unter Mitarbeit von Stephanie Graw</i>	
Wie Schulinspektor/innen fragen und Grundschüler/innen antworten	177

<i>Peter Schulze & Christiane von Schachtmeyer</i>	
Der Blick aus der Praxis – die Schulformexpertinnen und -experten	219

Teil 3: Die Befunde der Schulinspektion

<i>Melanie Ehren & Jaap Scheerens</i>	
Evidenzbasierte Referenzrahmen zur Schulqualität als Grundlage von Schulinspektion	233

<i>Holger Gärtner</i>	
Zusammenhang von Schul- und Unterrichtsqualität und schulischen Rahmenbedingungen	273

<i>Marcus Pietsch & Timo Ehmke</i>	
Die Qualität des Unterrichts an SINUS-Grundschulen. Eine Programmevaluation mithilfe von Daten der Schulinspektion Hamburg	295

<i>Klaudia Schulte & Markus Lücken</i>	
Der Einfluss schulischer Prozesse auf die Lernentwicklung der Schülerinnen und Schüler an weiterführenden Schulen in Hamburg	317

<i>Susanne Peters</i>	
Eltern als Stakeholder von Schule. Erkenntnisse über die Sichtweise von Eltern durch die Hamburger Schulinspektion.	341

Teil 4: Die Zukunft der Schulinspektion

<i>Sabine Müller</i>	
Zukunft der externen Evaluation – Trends und Perspektiven in Deutschland. .	369

<i>Hans-Jürgen Kuhn</i>	
Rechtliche, strukturelle und politische Rahmenbedingungen für externe Evaluation in Deutschland.	385

Martina Diedrich

Der zweite Zyklus der Schulinspektion Hamburg: ein Ausblick. 419

*Martin Spiewak, Angela Hellberg, Claudia Lemke, Gerrit Petrich,
Norbert Rosenboom, Mina Zulal*

Die Perspektive der Stakeholder: Ein Interview von Martin Spiewak 437

Autorinnen und Autoren 455

Vorwort

Vor rund zehn Jahren wurden die ersten Schulinspektorate in Deutschland eingeführt. Als Teil einer veränderten Steuerungsstrategie, die die Qualitätssicherung und -entwicklung von Schulen ins Zentrum setzt, sollen sie die Qualität schulischer Prozesse evaluieren, prozessuale Mindeststandards in Schule und Unterricht sichern und darüber hinaus Schulentwicklung stimulieren.

Die Hamburger Schulinspektion ist keine Inspektion der allerersten Stunde, wurde sie doch erst im Jahr 2007 offiziell eingerichtet. Zu diesem Zeitpunkt waren in vielen anderen Bundesländern schulische Evaluatorinnen und Evaluatoren bereits zwei bis drei Jahre im Feld und hatten die unterschiedlichsten Verfahren entwickelt und verschiedene Konzepte erprobt. Von ihren vielfältigen Erfahrungen mit diesem in Deutschland dahin unbekannten Instrument konnte die Hamburger Schulinspektion profitieren.

Gleichwohl hatte man in Hamburg bereits seit der Jahrtausendwende erste Konzepte zur externen Prozessevaluation an Einzelschulen in Hamburg diskutiert und entwickelt. Den Impuls hierfür gaben insbesondere die Ergebnisse der Längsschnittstudie LAU (Aspekte der Lernausgangslage und der Lernentwicklung), die seit Mitte der 1990er Jahre regelmäßig durchgeführt wurde. Es vergingen jedoch noch mehrere Jahre, bis die erste Hamburger Schule tatsächlich durch ein Inspektionssteam evaluiert wurde. Seitdem bewegt sich die Schulinspektion im Hamburger Schulsystem stets im Spannungsfeld von Administration, Praxis und Wissenschaft und muss dabei ihren eigenen sowie den Ansprüchen der verschiedenen schulischen und außerschulischen Stakeholder gerecht werden.

In ihrem ersten Zyklus, der im Jahr 2007 begann und bis zum Jahr 2013 andauerte, inspizierte die Schulinspektion 375 Hamburger Schulen. Während dieser sechs Jahre nahmen etwa 10.000 Lehrerinnen und Lehrer, gut 36.000 Eltern, rund 92.000 Schülerinnen und Schüler sowie etwa 2.000 Vertreterinnen und Vertreter Hamburger Unternehmen an den schriftlichen Befragungen der Inspektion teil. Darüber hinaus wurden fast 30.000 Unterrichtssequenzen durch die Inspektorinnen und Inspektoren beobachtet und etwa 2.000 Interviews mit Schulleitungen, Eltern- und Schülerräten sowie Lehrerkollegien und weiteren Mitarbeiterinnen und Mitarbeitern der inspizierten Schulen, aber auch mit Vertreterinnen und Vertretern von Ausbildungsbetrieben geführt.

Die Stärke des vorliegenden Bandes ist es, so meinen wir, dass er der besonderen Stellung der Schulinspektion im Gesamtgefüge des Hamburger Bildungssystems Rechnung trägt, indem er Akteure aus den verschiedenen Bereichen und mit unterschiedlichen Perspektiven zu Wort kommen lässt. Ein Teil der Beiträge ist empirisch ausgerichtet und analysiert die Arbeit der Inspektion oder nutzt die im Rahmen der Schulinspektion Hamburg erhobenen Daten für weiterführende

Analysen zum Hamburger Schulsystem. Andere Beiträge wiederum setzen sich mit den Rahmenbedingungen, Grundlagen und Folgen von Schulinspektion auseinander oder widmen sich den Erfahrungen und Erwartungen einzelner Akteure und Akteursgruppen.

Wir bedanken uns herzlich bei allen Autorinnen und Autoren dieses Bandes für die kooperative, kollegiale und anregende Zusammenarbeit. Insbesondere haben wir uns darüber gefreut, dass die vielen Kolleginnen und Kollegen aus diversen Kontexten und Arbeitsfeldern umstandslos bereit waren, ihre Expertise in Form fundierter und kritischer Beiträge sowie Gutachten beizusteuern. Sehr glücklich sind wir vor allem auch darüber, dass sich die Vertreterinnen und Vertreter der schulischen Stakeholder aktiv in diese Publikation eingebracht haben.

In der Zusammenarbeit mit all diesen verschiedenen Personen ist uns deutlich geworden, dass es institutionen- und akteursübergreifend ein großes Interesse daran gibt, über gute Bildung in und für Hamburg ins Gespräch zu kommen und sie gemeinsam zu gestalten. Dafür bieten – nicht nur, aber auch – die Schulinspektion und ihre Ergebnisse konstruktive Gesprächsanlässe. In diesem Sinne wünschen wir Ihnen eine ertragreiche Lektüre!

Hamburg im Juni 2015

Marcus Pietsch, Barbara Scholand & Klaudia Schulte

Expertenurteile – Achillesferse oder Trumpf der Schulinspektion?

Moritz G. Sowada

Zusammenfassung

Die Bewertung der Qualität von Schulen durch Inspektionsteams ist ein konstitutives Element von Schulinspektionsverfahren. Dieser Beitrag gibt einen Überblick über einschlägige Forschungsliteratur zu Qualitätsurteilen im Rahmen von Schulinspektion. Es wird dabei deutlich, dass erkenntnistheoretische Annahmen in die Forschung mit einfließen, die beeinflussen, ob Differenzen in den Urteilen von Inspektorinnen und Inspektoren per se als problematisch gelten. Anschließend wird gezeigt, welche innovativen Ansätze zur Erforschung von Urteilen, Entscheidungen und Bewertungen in der Psychologie und Soziologie verwandt werden und welchen Mehrwert deren Adaption für die Erforschung von Expertenurteilen in der Schulinspektion verspricht. Schließlich werden unterschiedliche Optionen der Professionalisierung von Expertenurteilen angedeutet.

1. Einleitung

Es ist die Aufgabe von Schulinspektionsteams, die Qualität von Schulen auf Basis von standardisierten Kriterien und Indikatoren zu bewerten. Für diese voraussetzungsvolle Aufgabe wird meist qualifiziertes Personal eingesetzt. So bestehen z. B. die Inspektionsteams in Hamburg normalerweise aus zwei hauptamtlichen Inspektoren bzw. Inspektorinnen, einem Leitungsmitglied einer anderen Schule der gleichen Schulform (sogenannte Schulformexperten bzw. Schulformexpertinnen) und – an beruflichen Schulen – noch einer weiteren Person als Vertretung der Wirtschaft (vgl. IfBQ 2011, S. 6; Döbert et al. 2008, S. 87–91).

Die fachliche Qualifikation und die berufliche Erfahrung der Inspektorinnen und Inspektoren sind bedeutsam. Sie stellen Ressourcen dar, die kompetente Expertenurteile ermöglichen. Die Berufserfahrung der Hamburger Inspektorinnen und Inspektoren wird wie folgt charakterisiert: „Von den 15 Personen, die Inspektionen hauptamtlich durchführten, haben neun ein Lehramtsstudium absolviert mit einer schulischen Unterrichtserfahrung zwischen 3 und 17 Jahren. Die übrigen Inspektorinnen und Inspektoren kommen aus geistes- und sozialwissenschaftlichen Berufen. Alle verbindet, dass sie über eine vertiefte Kenntnis sowohl von Schule als auch von Qualitätssicherung und Qualitätsentwicklung verfügen“ (IfBQ 2012,

S. 11). Sie erhalten zudem vor Übernahme von Teamleitungen eine etwa viermonatige Zusatzausbildung (vgl. IfBQ 2014). Hauptamtliche Inspektorinnen und Inspektoren erwerben darüber hinaus kontinuierlich berufspraktische Erfahrung in der Beurteilung von Schulen im Laufe ihrer Inspektionstätigkeit. Sie können daher als Expertinnen und Experten für die externe Evaluation von Schulen gelten.

Die Einschätzung und Bewertung von Schule und Unterricht anhand vorab definierter Qualitätsmerkmale ist der Kern der Expertise von Inspektorinnen und Inspektoren und steht damit im Zentrum ihrer beruflichen Tätigkeit. Trotz der zentralen Funktion des Bewertens im Prozess der Schulinspektion liegen vergleichsweise wenig gesicherte Erkenntnisse darüber vor, wie Inspektorinnen und Inspektoren Bewertungen vornehmen. Und dies, obwohl der bildungspolitischen Erwartung, dass Schulen aus den Inspektionsergebnissen Schul- und Unterrichtsentwicklungsmaßnahmen ableiten, diese durchführen und schließlich evaluieren sollen (vgl. Altrichter et al. 2014; Sommer 2011). So gibt es zwar präskriptive Vorgaben dazu, wie Schulinspektionen durchgeführt werden sollen, diese sind jedoch als Richtlinien zu verstehen (vgl. Nutley et al. 2012). Solche Bewertungsvorgaben sind unterdeterminiert, d. h. sie können nur in eingeschränktem Maße vorwegnehmen, wie Inspektorinnen und Inspektoren in der Praxis tatsächlich bewerten (vgl. Sadler 2009, 2014). Bewertungsrichtlinien lassen daher nur begrenzt Schlüsse auf tatsächliche Bewertungspraxen zu.

In der Regel vollziehen Inspektorinnen und Inspektoren ihre Bewertungsarbeit unter restriktiven Rahmenbedingungen. Sie müssen mit begrenzten Personalressourcen und unter Zeitdruck umfangreiche Daten erheben, analysieren und ggf. synthetisieren, um eine Vielzahl an Qualitätskriterien zeitnah zu bewerten (vgl. Gray & Wilcox 1995, S. 139). Wie Inspektorinnen und Inspektoren dies vollbringen, ist weitgehend ungeklärt. Dennoch wird die Expertise von Inspektorinnen und Inspektoren von Boyne, Day und Walker (2002) als zentrale Variable für die Kosten-Nutzen-Relation von Inspektionen postuliert – ohne Expertise jedoch inhaltlich zu bestimmen. Sie operieren mit folgender Annahme: Je höher die Expertise der Inspektorinnen und Inspektoren ist, desto größer ist der Nutzen von Inspektionen und desto geringer sind die damit verbundenen direkten und indirekten Kosten.

Anstatt den Prozess des Bewertens zu untersuchen, hat die empirische Bildungsforschung sich dem Forschungsfeld Schulinspektion bisher hauptsächlich über die Wirkungsfrage genähert (vgl. Dederling 2012; Husfeldt 2011; Kotthoff & Böttcher 2010). Es wird untersucht, was im Zusammenhang mit Schulinspektion an Schulen passiert. Dabei wird oft die Sicht der Schulleitung und manchmal auch der Lehrkräfte abgefragt (vgl. Lambrecht & Rürup 2012, S. 65). Teilweise wird zwischen intendierten und nicht intendierten sowie erwünschten und unerwünschten Wirkungen unterschieden (vgl. Ehren & Visscher 2006). Auch international gibt es kaum Wissen dazu, was eine „wirksame“ Schulinspektion ausmacht. Es ist jedoch bekannt, dass groß angelegte Reformmaßnahmen im Bildungswesen ca.

sieben Jahre bedürfen, bevor entsprechende Wirkungen verlässlich nachweisbar sind (vgl. Borman et al. 2003). Die vorgelagerte Frage, wie Schulinspektionsergebnisse zustande kommen, wird hingegen in der empirischen Forschung weitgehend ausgeblendet (vgl. Lambrecht & Rürup 2012, S. 72). Dieses Desiderat ist Gegenstand des vorliegenden Textes.

Es gilt grundsätzlich zu beachten, dass mit Schulinspektion unterschiedliche Funktionen erfüllt werden sollen (vgl. Maritzen in diesem Band). Diese Funktionen können in einem Spannungsverhältnis zueinander stehen, so z. B. die Kontroll- und Entwicklungsfunktion (vgl. Böttcher & Keune 2012). Dieses Spannungsverhältnis äußert sich für die Inspektorinnen und Inspektoren in einem „Rollenkonflikt zwischen objektivem Beobachter und Akteur, der über schulisches Wissen verfügt, aber nicht beraten darf“ (Preuß et al. 2015, S. 136). Je nachdem, ob die Kontroll- oder Entwicklungsfunktion hervorgehoben wird, sind unterschiedliche Entwicklungsstrategien zu favorisieren (vgl. hierzu Abschnitt 5).

Im Folgenden werden einschlägige Forschungsbefunde zu Bewertungen und Bewertungsprozessen von Inspektorinnen und Inspektoren vorgestellt. Zuerst werden kritische Anfragen hinsichtlich der Methodologie und Methoden von Schulinspektion referiert. Im Anschluss daran wird auf Ergebnisse aus Inter-Beurteiler-Studien eingegangen und diese daraufhin um Befunde qualitativer Forschung ergänzt. Im dritten Abschnitt werden Heuristiken als vielversprechende Option zur Erforschung von Schulinspektionsurteilen eingeführt. Kontrastierend dazu wird im vierten Abschnitt die Analyse von Wertordnungen thematisiert, die Bewertungsdifferenzen zugrunde liegen können. Im fünften und letzten Abschnitt werden die Potenziale der angerissenen Ansätze für zukünftige Bewertungsforschung im Kontext von Schulinspektion erörtert.

2. Urteile von Inspektorinnen und Inspektoren

An dieser Stelle erfolgt eine knappe Bestandsaufnahme der Literatur, die bisher die Bewertung von Inspektorinnen und Inspektoren explizit zum Gegenstand gemacht hat. Es gibt zum einen psychometrische Ansätze, in denen die Güte der vorgenommenen Bewertungen von Inspektorinnen und Inspektoren Analysegegenstand ist. Geprüft wird mithilfe statistischer Verfahren, inwieweit klassische wissenschaftliche Gütekriterien wie Objektivität, Reliabilität und Validität erfüllt werden. Entsprechende Erhebungen basieren meist auf der Bewertung ein und derselben Unterrichtssequenz durch mehrere Inspektorinnen und Inspektoren auf Basis eines standardisierten Unterrichtsbeobachtungsbogens. Die Annahme ist zumeist, dass je größer die Übereinstimmung zwischen den bewertenden Personen ist, desto verlässlicher sei das Bewertungsergebnis. Die zugrunde liegende erkenntnistheoretische Prämisse beinhaltet, dass es eine – und nur eine – korrekte Bewertung

gibt. Abweichungen von diesem hypothetisch „wahren“ Wert seien daher als problematisch einzustufen und sollten dementsprechend soweit wie möglich reduziert werden.

Andererseits gibt es qualitativ orientierte Forschungsberichte, die grundlegend infrage stellen, ob eine umfassende Standardisierung des Bewertungsverhaltens von Inspektorinnen und Inspektoren erstrebenswert oder gar möglich ist. Gerade bei so komplexen Phänomenen wie Schul- und Unterrichtsqualität gäbe es Ermessensspielräume, deren Vereindeutigung sachunangemessen sei. Zur Stützung dieser Position wird auf unterschiedliche Argumente zurückgegriffen, die unten weiter ausgeführt werden. Zuvor werden jedoch empirische Befunde aus dem psychometrischen Forschungsansatz angeführt.

2.1 Psychometrische Forschung

Wie erwähnt zielt psychometrische Forschung u. a. darauf ab, zu überprüfen, in welchem Maße die klassischen Gütekriterien seitens des Inspektionsverfahrens abgesichert und durch Inspektorinnen und Inspektoren erfüllt werden. Dies impliziert die Annahme, dass solche wissenschaftlichen Gütekriterien gegenstandsangemessen sind. Gärtner (2013) attestiert Schulinspektionssystemen in diesem Zusammenhang eine „positivistische [...] Grundorientierung“ (S. 698). Schulinspektion selbst wird – allerdings in unterschiedlichem Maße – an dem Ideal der evidenz- bzw. wissenschaftsbasierten Evaluation von Schulen ausgerichtet (vgl. Rürup 2008). Es sollte daher nicht verwundern, wenn Schulinspektion an diesen Ansprüchen gemessen und in der Folge dann kritisiert wird. Andererseits identifiziert Rürup (2008) neben einer wissenschaftlichen Orientierung von Schulinspektionssystemen auch eine konkurrierende, da den *objektiven* Vergleich erschwerende Einzelfallorientierung. Das Verhältnis von Standardisierung und Fallorientierung fällt dabei zwischen den Inspektionsverfahren in den einzelnen Bundesländern unterschiedlich aus.

2.1.1 Methodisch-methodologische Kritik an Schulinspektionsverfahren

In England sind mehrere Aufsätze veröffentlicht worden, die kritische Fragen bezüglich der methodischen bzw. methodologischen Güte von Schulinspektion formulieren (vgl. Fitz-Gibbon & Stephenson-Forster 1999; Gray & Wilcox 1995) und dabei z. T. Forschungsstandards als Referenzpunkt heranziehen. Ähnlich gelagerte Kritik wird auch gegenüber der Inspektion der ersten Phase der Lehrerbildung (*Initial Teacher Training*) an Hochschulen in England geäußert (vgl. Cale & Harris 2003; Campbell & Husbands 2000; Sinkinson & Jones 2001). Problematisiert wird

u. a. die Auswahl von zu beobachtenden Unterrichtssequenzen und damit deren Repräsentativität für die Unterrichtsqualität an einer Schule (vgl. Wilcox & Gray 1996, S. 70 f.). Ebenfalls wird angezweifelt, ob Schulinspektion die Reliabilität und Validität u. a. von Unterrichtsbewertungen gewährleisten kann (vgl. Fitz-Gibbon & Stephenson-Forster 1999).

Auch im deutschsprachigen Kontext werden die methodisch-methodologischen Herausforderungen, insbesondere im Zusammenhang mit Unterrichtsbewertungen, thematisiert. So wird etwa der mit 20 Minuten relativ kurze Beobachtungszeitraum für die Unterrichtsbewertungen problematisiert (vgl. Bos et al. 2006). Zu dieser begrenzten Beobachtungsdauer kommt hinzu, dass Inspektorinnen und Inspektoren mitunter relevante Kontextinformationen fehlen und die Unterrichtsbewertung selbst die Beobachtungssituation verfälschen könnte (vgl. Praetorius 2013). Nicht zuletzt werden Beurteilereffekte als Gefährdung der Aussagekraft des Ergebnisses sowie deren Nutzens angesehen (vgl. Pietsch & Tosana 2008).

2.1.2 *Inter-Beurteiler-Studien zur Unterrichtsbewertung*

Die Studie von Matthews, Holmes, Vickers und Corporaal (1998) kann als Reaktion auf die methodisch-methodologische Kritik an dem Verfahren des englischen Schulinspektorats Ofsted (*Office for Standards in Education*) gedeutet werden. Sie haben die Reliabilität und Validität von Unterrichtsbewertung durch jeweils zwei Inspektorinnen bzw. Inspektoren untersucht. Die freiwillig teilnehmenden Inspektorinnen und Inspektoren haben ihre Bewertung von 173 gemeinsam beobachteten Unterrichtssequenzen getrennt voneinander festgehalten. Es wurde für jede Unterrichtssequenz jeweils eine Bewertungskategorie von 1 bis 7 vergeben. In 66 Prozent aller Fälle wurde von dem jeweiligen Bewertungspaar die gleiche Bewertungskategorie gewählt (vgl. Matthews et al. 1998, S. 182). Somit besteht ein Bewertungsdissens in einem Drittel aller untersuchten Fälle, der allerdings im kritischen Grenzbereich von *satisfactory* zu *unsatisfactory* (Bewertungsstufen 4 und 5) höher ausfällt (vgl. ebd., S. 184).

Auch für Deutschland liegen erste Inter-Beurteiler-Studien mit Inspektorinnen und Inspektoren vor. Pietsch und Tosana (2008) konnten in ihrer Hamburger Studie feststellen, dass „nur geringe Urteilsfehler nachweisbar sind (rund 9 Prozent der Gesamtvariation), wobei diese vor allem auf die unterschiedliche Interpretation von Items durch die bewertenden Unterrichtsbeobachter zurückzuführen sind“ (S. 449). Sie verweisen aber auch darauf, dass die Inspektorinnen und Inspektoren bei der Unterrichtsbewertung unterschiedliche Anspruchsniveaus anlegen, d. h., einige urteilen strenger als andere. In einer Nachfolgestudie wurde jedoch herausgestellt, dass diese Strengeeffekte einen geringeren Anteil an der Beurteilervarianz haben als die variierende Interpretation der Items des Unterrichtsbewertungs-

bogens durch die Inspektorinnen und Inspektoren (vgl. Pietsch 2010, S. 128). In einer Vergleichsstudie haben Müller und Pietsch (2011) Inspektorinnen und Inspektoren aus sechs Bundesländern videografierte Unterrichtssequenzen mit verschiedenen Unterrichtsbewertungsbögen beurteilen lassen. Sie stellten dabei eine zum Teil sehr hohe Beurteilervarianz von bis zu 71 Prozent fest (vgl. Müller & Pietsch 2011, S. 45). Dennoch seien aggregierte Aussagen zur Unterrichtsqualität auf Schulebene, laut den Ergebnissen der Studie, statistisch zulässig (vgl. ebd., S. 51).

Die in der Literatur aufgeworfene methodisch-methodologische Kritik wurde also im Rahmen von Inter-Beurteiler-Studien adressiert. Diese empirischen Studien weisen Beurteilervarianz nach, wobei das Ausmaß der den Inspektorinnen und Inspektoren zugerechneten Varianz selbst eine große Spannbreite aufweist. Über die Gründe für die festgestellte Beurteilervarianz können die Studien hingegen keine Auskunft geben, da der Prozess der Beurteilung und die Gründe der Inspektorinnen und Inspektoren für die Wahl einer Bewertungskategorie methodisch ausgeblendet werden. An dieser Stelle können Interview-, *Think-Aloud*- und ethnographische Studien weiterführen. Auf entsprechende qualitative Forschungsstränge wird daher nun auch eingegangen.

2.2 Qualitativ-orientierte Forschung

Qualitative Studien, die die Bewertungsarbeit von Inspektorinnen und Inspektoren in den Blick nehmen, sind relativ dünn gesät. Entsprechende Studien haben zu meist ein dezidiert *nicht* positivistisches Verständnis zugrunde gelegt (vgl. Lincoln et al. 2011). Dies hat Auswirkungen darauf, wie Schulinspektion als Forschungsgegenstand verstanden wird. So ist das Forschungsinteresse nicht darauf gerichtet, wie etwa die „tatsächliche Qualität“ von Schule und Unterricht mit optimierten Instrumenten und Verfahren besser erfasst werden kann. Dies deswegen, weil in entsprechenden Paradigmen nicht von *einer* singulären, objektiv erfassbaren Qualität ausgegangen wird. Stattdessen liege Qualität im Auge des Betrachters, ohne jedoch deswegen beliebig zu sein (vgl. Heid 2013). Vorstellungen von guter Schule und gutem Unterricht sind in soziohistorische Kontexte eingebunden, die Qualitätszuschreibungen stabilisieren. Qualität ist in einer solchen Perspektive Aushandlungsgegenstand zwischen den Stakeholdern. Wenn es also nicht darum geht, der *einen* objektiven Qualität so nah wie möglich zu kommen, worum geht es dann in qualitativer Forschung zu Bewertungen im Kontext von Schulinspektion? Um hierauf eine Antwort zu geben, werden ausgewählte qualitative Studien nachfolgend skizziert.

Interviewstudien sind eine häufige Variante qualitativer Schulinspektionsforschung, da sie vergleichsweise einfach zu realisieren sind. In Interviewstudien ge-

ben Inspektorinnen und Inspektoren selbst Auskunft zu ihrer Bewertungstätigkeit. Sie können dabei als Expertinnen und Experten verstanden werden, die direkt und informativ zu ihren Bewertungspraxen relevante Informationen angeben können. Andererseits kann im Reden über Schulinspektion mehr mitgeteilt werden, als den Akteuren in der Situation unbedingt bewusst ist.

2.2.1 *Expertenurteile als Professional Judgements*

Nixon und Rudduck (1993) beschäftigt die Frage, wie Expertenurteile (*Professional Judgements*) aus Sicht von Inspektorinnen und Inspektoren zustande kommen. Die Einführung neuer Schulinspektionsverfahren unter Ofsted wurde als Wandel von impliziten zu expliziten Bewertungskriterien beschrieben. Für Nixon und Rudduck (ebd.) greift diese Charakterisierung zu kurz. Sie verweisen zum einen in einer philosophisch gestützten Argumentation darauf, dass die explizit benannten Kriterien den Bewertungsprozess zwar anleiten, aber nicht determinieren können. Die Kriterien haben vielmehr eine orientierende Funktion für einzelne Inspektorinnen und Inspektoren, Inspektionsteams und Schulen (vgl. ebd., S. 140 f.).

Ihre Interviewbefunde verweisen auf eine Differenz zwischen den Selbstaussagen der Inspektorinnen und Inspektoren zu ihren Urteilspraxen und einer kriterienbasierten Bewertungslogik des offiziellen Schulinspektionsverfahrens. Inspektorinnen und Inspektoren würden Such- und Bewertungsprozesse unter Zuhilfenahme ihres impliziten Wissens bewerkstelligen (vgl. ebd., S. 143). Dabei verschiebt sich der Fokus der Aufmerksamkeit situationsspezifisch, z. B. von der Erfüllung von binär kodierten Vorschriften hin zur Einschätzung von Qualität im Gesamtzusammenhang (vgl. ebd., S. 144). Hieran thematisch anschließend besteht für Meyer (2012, S. 46–49) die Inspektionstätigkeit nicht allein aus der Addition von beobachteten Fakten zu einer Gesamtbewertung, auch die Wertung einzelner Elemente basiere auf der Gesamtbetrachtung einer Schule. Nur so sei eine angemessene Bewertung von komplexen Qualitätsaspekten überhaupt möglich. Expertenurteile von Inspektorinnen und Inspektoren beruhen demnach nicht nur auf der Bewertung von voneinander isolierten Fakten, sondern gerade bei komplexeren Bewertungsobjekten auf einem hermeneutischen Bewertungsprozess.

Wilson (1996) verweist auf weitere Aspekte, die die Komplexität von Expertenurteilen verdeutlichen. Seine zehnmonatige Feldstudie aus dem Jahre 1992 fällt in eine Zeit, in der das englische Schulinspektionswesen tief greifend umstrukturiert wurde. Die Reform kann als eine Umstellung von einem *Connaissanceur*-Modell, das auf Expertise und Kennerschaft beruhte (vgl. Eisner 2013), hin zu einem sozialwissenschaftlichen Inspektionsverfahren, das auf Standardisierung und expliziten Kriterien basiert, verstanden werden (vgl. Scheerens et al. 2003, S. 44 f.). Gleichwohl, so Wilson (1996, S. 122), seien Expertenurteile auch weiterhin erfahrungs-

gebunden und abhängig von unbewussten kognitiven Prozessen. Die Gründe für ein Expertenurteil sind daher nur begrenzt explizierbar. Gerade das Praxiswissen (*Practitioner Knowledge*) von Inspektorinnen und Inspektoren ermögliche kompetente Expertenurteile. Kriterienkataloge sind zwar standardisiert, aber deren fallbezogene Anwendung erfordere ein gewisses Maß an Intuition und Expertise (vgl. Nutley et al. 2012, S. 878). Dass dies potenziell zu unterschiedlichen Urteilen führen könne, sei keine Schwäche, sondern eine Stärke von Schulinspektion, so Wilson (1996). In dem Verfahren ginge es nicht darum, absolute, sondern nur relationale Urteile zu fällen, die zur Auseinandersetzung mit ihnen innerhalb einer Schulgemeinschaft anregen sollten. Wilson (ebd.) betont primär die Entwicklungsfunktion von Schulinspektion.

Inspektorinnen und Inspektoren seien darum bemüht, ihre Qualitätseinschätzungen mit Evidenzen abzusichern und ihnen dadurch Gewicht zu verleihen. Das Verhältnis von Evidenz und Expertenurteil sei von Wechselwirkungen geprägt. Es werden demnach nicht erst Evidenzen systematisch gesammelt und dann abschließend beurteilt, sondern Bewertung und Datensammlung finden während der gesamten Schulinspektion parallel statt. So beeinflussen vorgängige Einschätzungen von Inspektorinnen und Inspektoren, welche Daten zu Evidenzen werden (vgl. Wilson 1996, S. 205). Dabei gehe es darum, die vorläufigen, aber dennoch aufmerksamkeitsstrukturierenden Einschätzungen und Annahmen mit Evidenzen abzusichern. Die Suche nach konträren Evidenzen kommt hingegen seltener zum Einsatz, da sie eine aufwendigere Prüfmethode darstellt (vgl. Nutley et al. 2012, S. 878).

Dass diese Erkenntnisse vor dem Hintergrund eines positivistischen Paradigmas problematisch sind, zeigen folgende Studien.

2.2.2 Expertenurteile zwischen Programmatik und Pragmatik

Millett und Johnson (1998) verweisen darauf, dass Ofsted von den Inspektorinnen und Inspektoren erwartet, dass sie die Effektivität von Unterricht bewerten und dabei ihre persönlichen Präferenzen für bestimmte Methoden und Unterrichtsstile außen vor lassen. Etwaige Präferenzen für einen bestimmten pädagogischen Stil werden dementsprechend abgewertet, wohingegen fachliche Expertise in z.B. Mathematik im Bewertungsprozess ausdrücklich gewünscht sei. Solche politisch motivierten Vorgaben operieren mit der Prämisse, dass es möglich sei, bestimmte Formen von Vorwissen zu exkludieren, andere hingegen zu inkludieren. Die interviewten Inspektorinnen und Inspektoren geben an, dass sie versuchen, solchen Vorgaben zu entsprechen. Gleichwohl scheint es, dass sie dies auf unterschiedliche Art und Weise tun. Solche abstrakten Vorgaben werden rekontextualisiert, d. h. mit unterschiedlichem Sinn versehen und darüber in das eigene Aufmerksamkeits- und Urteilsverhalten integriert (vgl. Fend 2008). Mathematisch stärker vorgebildete In-

spektorinnen und Inspektoren bewerten, so mutmaßen Millett und Johnson (1998), eher die fachliche Qualität des Unterrichts, wohingegen andere stärker die allgemeine Unterrichtsqualität in den Blick nehmen.

Auch Lindgren (2015) nimmt das Verhältnis von Bewertungsvorgaben und Urteilspraxen in den Blick. Er benutzt hierfür Goffmans konzeptuelle Unterscheidung zwischen Vorder- und Hinterbühne (Goffman 1959). Auf der Vorderbühne werde von Inspektionsteams angestrebt, einheitlich gegenüber der Schulöffentlichkeit aufzutreten. Es werde eine konsistente Ergebnisdarstellung angestrebt. Dabei gelte es, die formale Übereinstimmung mit den offiziellen Verfahrensvorgaben sicherzustellen. Zum anderen werde darauf hingewirkt, das Ergebnis einheitlich als Team zu vertreten. Bevor ein Schulinspektionsbericht als Produkt auf der Vorderbühne bekannt gegeben wird, wird auf der Hinterbühne auf vielfältige Art und Weise darauf hingearbeitet. Lindgren kann zeigen, dass auf der Hinterbühne durchaus methodologische Unsicherheiten, z. B. bzgl. der Herstellung von Urteilskonsistenz, vorherrschen. Solche Interna werden gegenüber Außenstehenden jedoch nicht kommuniziert. Divergente Einschätzungen von Qualitätsaspekten werden vielmehr in geschützten Gesprächen vereinheitlicht. Zudem werden negative Bewertungen eingehender geprüft und aufwendiger abgesichert als positive Urteile (siehe auch Nutley et al. 2012, S. 879). Dabei gehe es *nicht nur* um das richtige, da evidenzbasierte Expertenurteil, sondern auch um Fragen der antizipierten Ergebnisrezeption aufseiten der Schule.

2.3 Zwischenfazit

Die angeführten quantitativen und qualitativen Studien zeigen, dass Expertenurteile in sehr unterschiedlicher Weise zum Gegenstand von Forschung gemacht werden. Wo quantitative Studien eine Komplexitätsreduktion vornehmen, indem mittels Inter-Beurteiler-Studien Fragen der Konsistenz bzw. Variation von Bewertungen unter Absehung des Bewertungsprozesses untersucht werden, wird in qualitativen Studien oftmals Komplexität gesteigert. Expertenurteile werden dabei nicht allein als technisch-methodische, sondern als mikropolitisch machtbasierte Vorgänge konzeptualisiert. Durch diese andere Perspektive kommen Aspekte jenseits der Konsistenzfrage in den Blick.

Die bisherige Forschung zum Thema, auf welche Weise und wie konsistent Inspektorinnen und Inspektoren bewerten, ist erst in den Anfängen begriffen. Da evidenzbasierte Expertenurteile gleichwohl den Kern der Arbeit von Inspektorinnen und Inspektoren darstellen, werden nachfolgend vielversprechende Forschungsansätze in benachbarten Forschungsfeldern vorgestellt, die dazu geeignet sind, zukünftige Forschung zu Schulinspektionsurteilen in neue und innovative Richtungen zu führen.

3. Psychologie des Urteilens und Entscheidens

In den letzten Jahrzehnten sind bedeutsame Fortschritte in der psychologischen Urteils- und Entscheidungsforschung erzielt worden. Ausgangspunkt war die von Simon (2000) Mitte der 1950er popularisierte Einsicht, dass Menschen nur begrenzt rational (*Bounded Rationality*) handeln können (vgl. Gigerenzer & Selten 2002). Aufgrund begrenzter kognitiver Kapazitäten, endlicher Zeit und unzuverlässiger Informationen in der Umwelt ist es Menschen nicht möglich, optimale rationale Urteile und Entscheidungen zu treffen. Diese limitierenden Bedingungen gelten auch für die Bewertungsarbeit von Inspektorinnen und Inspektoren. Hierzu liegen jedoch bisher keine empirischen Forschungsergebnisse vor.

In der Psychologie gibt es hingegen eine Vielzahl an Forschungsbefunden, die zeigen, dass Menschen bei ihren Urteilen und Entscheidungen Heuristiken verwenden (vgl. Gigerenzer & Gaissmaier 2011). Heuristiken sind kognitive Prozesse, mit denen Menschen viable Urteile und Entscheidungen *unter Unsicherheit* treffen. Heuristiken können bewusst oder unbewusst ablaufen. Ein Vorteil von Heuristiken ist, dass sie in komplexen Situationen effiziente Urteils- und Entscheidungsfindung ermöglichen, indem gerade nicht alle verfügbaren Informationen berücksichtigt werden. Andererseits können diese Abkürzungsstrategien potenziell zu systematischen Urteilsverzerrungen führen (vgl. Tversky & Kahneman 1974). Es ist zu erwarten, dass Inspektorinnen und Inspektoren Heuristiken bereichsspezifisch einsetzen oder gar aufgabenspezifische Heuristiken verwenden.

Es gibt in der Psychologie des Urteilens und Entscheidens zwei Forschungsstränge, in denen Heuristiken bedeutsam sind, die jedoch mit unterschiedlichen Prämissen und Forschungsvorgehen arbeiten (vgl. Hammond 1996). Sie werden nachfolgend vorgestellt.

3.1 *Judgment and Decision-Making*

Der Ansatz des *Judgment and Decision-Making* (JDM) nutzt vermeintlich einfache Aufgabenstellungen zum Vergleich menschlicher Urteile und Entscheidungen mit „korrekten“ Urteilen und Entscheidungen, die mittels formaler Logik abgeleitet werden und somit dem Ideal der Rationalität entsprechen. Rationalität wird diesem Ansatz gemäß als Kohärenz zwischen logischen und empirischen Urteilen und Entscheidungen gefasst. In diesem Ansatz erscheinen menschliche Urteile und Entscheidungen als defizitär und imperfekt, da postulierte Rationalitätsmaßstäbe insofern verletzt werden, dass tatsächliches Urteils- und Entscheidungsverhalten unter bestimmten reproduzierbaren Umständen von den abgeleiteten „rationalen Urteilen“ abweichen. Es handelt sich hierbei daher um einen präskriptiv-norma-

tiven Ansatz, der einen Beitrag dazu leisten soll, Entscheidungen und Urteile auf eine rationalere Basis zu stellen, um sie auf diese Weise zu optimieren (vgl. Kundin 2010, S. 349).

Es wird mittlerweile in einer Vielzahl kognitions- und sozialpsychologischer Modelle davon ausgegangen, dass das menschliche Denken auf zwei separaten Prozessen beruht (vgl. Evans 2008). Zum einen wird angenommen, dass es schnelle und weitgehend unbewusste Kognitionen gibt, die u. a. auf implizitem Wissen und impliziten Erfahrungen beruhen (System 1). Zum anderen gebe es langsamere, dafür aber bewusster ablaufende Kognitionen, die auf explizitem Wissen basieren (System 2). Für Expertenurteile von Inspektorinnen und Inspektoren ist daraus abzuleiten, dass beide Systeme in die Bewertungsarbeit mit einfließen. Das Verhältnis von beiden Systemen bei der Auswahl von Bewertungskategorien ist jedoch ungeklärt. Es ist aber davon auszugehen, dass Inspektorinnen und Inspektoren, um effiziente Bewertungsarbeit zu leisten, Kognitionen des Systems 1 bedürfen. Dieser Form der Kognition wird im folgenden Ansatz eine größere funktionale Bedeutung zugesprochen, als dies im JDM-Ansatz der Fall ist.

3.2 *Naturalistic Decision-Making*

Dieser zweite Ansatz fokussiert u. a., welche Rolle Expertise in realen Situationen spielt. Dieser Ansatz wird daher *Naturalistic Decision-Making* (NDM) genannt (vgl. Klein 2008). Hier interessiert u. a., inwiefern Experten sich von Novizen in ihrem Urteils- und Entscheidungsverhalten unter natürlichen Bedingungen unterscheiden. In einer mittlerweile klassischen Studie wurde untersucht, wie Einsatzleiter bei der Feuerwehr ihre Entscheidungen treffen (vgl. Klein et al. 2010). Die Studie zeigte, dass im Ernstfall nicht mehrere Handlungsalternativen entworfen und gegeneinander abgewogen werden, um die beste Handlungsoption zu identifizieren. Stattdessen haben die Einsatzleiter nur einen Lösungsansatz entworfen und bei positiver Ad-hoc-Evaluation umgesetzt.

Der NDM-Ansatz ist für die Forschung zu Urteilen von Inspektorinnen und Inspektoren insofern interessant, als er danach fragt, wie Akteure unter Bedingungen begrenzter Rationalität dennoch situationsadäquate Urteile und Entscheidungen zu treffen vermögen. Ebenfalls von Interesse ist, welche impliziten Strategien dabei verwandt werden. Der Rückgriff auf Feldstudien kann dabei die ökologische Validität der Befunde gegenüber dem Berufsalltag enthobenen Experimenten des JDM-Ansatzes steigern.

Ein weiterer Aspekt mit Relevanz für Schulinspektion ist die Frage danach, ob Expertinnen und Experten andere Bewertungsstrategien als Novizen erwerben und verwenden. Tourmen (2009) zeigt etwa, dass Berufserfahrung einen Einfluss darauf hat, welche konzeptuellen Ressourcen Evaluatorinnen und Evaluatoren bei ihrer

Bewertungsarbeit nutzen. Sie berichtet, dass Novizinnen und Novizen sich stark an methodischen Vorschriften orientieren, wohingegen Expertinnen und Experten eine stärkere Ergebnisorientierung aufweisen. Letztere würden auch stärker die soziale und politische Dimension von Evaluationen berücksichtigen und mögliche Evaluationsverläufe antizipieren. Sie empfiehlt, in der Ausbildung den Erwerb von konzeptuellem Denken stärker als bisher gegenüber Methodentraining zu fördern. Eine Übertragbarkeit entsprechender Befunde auf Inspektionen ist jedoch aufgrund von Unterschieden nur partiell möglich, denn es werden z. B. statt spezifischer Programme Schulen evaluiert. Darüber hinaus kommen bei der Schulinspektion standardisierte Instrumente zum Einsatz, wohingegen bei der Programmevaluation die Evaluation programmspezifisch konstruiert wird.

Die gesellschaftliche Einbettung von Expertenurteilen wird im folgenden Abschnitt thematisiert.

4. Soziologie der Bewertung

Es handelt sich hierbei um ein in der Entstehung begriffenes Forschungsfeld (vgl. Lamont 2012). Anders als die Psychologie des Urteilens und Entscheidens fokussiert es weniger auf Bewertung als psychisches, sondern als soziales und gesellschaftliches Phänomen. Entsprechende Zeitdiagnosen proklamieren, dass wir in einem Zeitalter der Rechenschaftspflicht (*Age of Accountability*) und in einer Audit-Gesellschaft bzw. Audit-Kultur leben (vgl. Power 1999; Strathern 2000). Schulinspektion kann daher als ein dem Zeitgeist entsprechendes Instrument der Rechenschaftslegung und Qualitätskontrolle verstanden werden.

Aus einer solchen soziologischen Perspektive ist in den Qualitätsrahmen und Kriterienkatalogen eine wertbasierte Hierarchie eingeschrieben und wird darüber zugleich festgeschrieben. Moderne Kommunikationstechnologien ermöglichen sodann die Dissemination solcher konzipierter Werthierarchien an die Schulen im Land und können dadurch quasi auf Distanz wirksam werden – noch bevor Inspektionsteams dort eintreffen (vgl. Prior 2003, S. 168).

Akteure und Akteursgruppen können zu dieser an sie herangetragenen Werthierarchie Stellung beziehen und werden dies unterschiedlich tun, je nachdem, ob sie damit übereinstimmende oder davon abweichende Wertvorstellungen haben. Konflikte im Rahmen von Schulinspektion können u. a. auf unterschiedlichen Werten beruhen. Ein für Schulinspektion charakteristisches Spannungsverhältnis besteht zwischen dessen Kontroll- und Entwicklungsfunktion (vgl. Böttcher & Keune 2012). Dieses Spannungsverhältnis impliziert konkurrierende Wertordnungen. Was ist wichtiger: Vergleichbarkeit zum Zwecke des System-Monitorings oder fallbezogene Förderung von Schulentwicklung? Nicht nur auf Verfahrensebene gibt es unterschiedliche Präferenzen, sondern auch bei den Inspektorinnen und Inspekto-

ren. Einige Inspektorinnen und Inspektoren wollen lediglich evaluieren, wohingegen andere die schulischen Akteure auch gerne beraten möchten.

Wenn es innerhalb eines Inspektionsteams zu unterschiedlichen Bewertungspräferenzen kommt, dann können die Teammitglieder versuchen, die anderen von ihren guten Gründen zu überzeugen (vgl. Sowada & Dederich 2014). Diese Rechtfertigungsarbeit (*Justification Work*) beinhaltet die Verhandlung des Werts einer Sache, wobei bestimmte gesellschaftlich präformierte Wertordnungen hervorgehoben und andere abgeblendet werden (vgl. Jagd 2011, S. 346). Expertenurteile sind demnach nicht neutral, sondern wertgebunden.

5. Diskussion und Ausblick

Lambrecht und Rürup (2012, S. 72) stellten fest, dass „der eigentliche Prozess des Inspizierens, aber auch der politische Kontext des Steuerungsinstruments merkwürdig unterbelichtet“ sei. Genau hier können Forschungsansätze aus den erziehungswissenschaftlichen Nachbardisziplinen neue Perspektiven jenseits der Wirkungs- und Wirksamkeitsfrage eröffnen.

Inspektorinnen und Inspektoren agieren – wie alle Menschen – unter Bedingungen begrenzter Rationalität. Die Konsequenzen für die Bewertungsarbeit von Inspektorinnen und Inspektoren sind bisher unzureichend erforscht. Gleichwohl wären Forschungsergebnisse zu den von Inspektorinnen und Inspektoren eingesetzten Heuristiken in vielerlei Hinsicht hilfreich. Sie könnten zur Weiterentwicklung von Schulinspektion beitragen, z.B. dadurch, dass entsprechendes Wissen ein aufgeklärteres Bewertungsverhalten fördert. Präskriptive Bewertungsvorgaben würden durch Wissen um Bewertungspraxen komplementiert. Entsprechende Erkenntnisse könnten dann in die Aus- und Weiterbildung von Inspektorinnen und Inspektoren integriert werden. Wenn bspw. Abkürzungsstrategien gefunden werden, die eine effizientere Bewertungsarbeit ermöglichen, ohne die erwünschte Qualität der Bewertung zu beeinträchtigen, dann würde durch deren Dissemination auch ein Beitrag zur Belastungsreduktion und zum Gesundheitsschutz geleistet werden.

Für ein besseres Verständnis der Bewertungspraxen von Inspektorinnen und Inspektoren sind interdisziplinäre Forschungsprojekte mit Fachleuten aus der Urteilsforschung angezeigt. Ziel sollte es sein, die eingesetzten Heuristiken zu explizieren. Dies kann in naturalistischen Settings, aber ggf. auch über den Einsatz von Fallvignetten im Rahmen von *Think-Aloud*-Studien geschehen. Erforderlich ist ein Wechsel von der bisher überwiegenden Wirksamkeits- und Wirkungsforschung hin zur Untersuchung von Bewertungsprozessen im Kontext von Schulinspektion. Für ein solches Forschungsprogramm kommt die Sekundärnutzung von bereits vorliegenden Inspektionsdaten leider nicht infrage, da der hier interessierende Prozess

der Generierung von Expertenurteilen darin nicht vorkommt. Wenn hingegen die offizielle Begründung von Expertenurteilen interessiert, so könnte die Analyse von Argumenten in Inspektionsberichten weiterführend sein.

Die Soziologie der Bewertung sensibilisiert dafür, dass das, was unter Qualität verstanden wird, historisch und kulturell kontingent ist. Daran anschließende Forschungsprojekte können z. B. die den Schulinspektionsverfahren zugrunde liegenden Wertordnungen explizieren. Dies kann helfen zu klären, ob wir inspizieren, was wir wertschätzen, oder wertschätzen, was wir inspizieren können (vgl. Biesta 2010). Forschung kann zudem untersuchen, wie unterschiedliche Akteure und Akteursgruppen darüber verhandeln oder gar darum ringen, was guter Unterricht und was gute Schule sei. Die in Qualitätsrahmen und Kriterienkatalogen kodifizierten Bestimmungsversuche sind nicht alternativlos und können hinterfragt werden – genauso wie die Ergebnisse einer Schulinspektion (vgl. Gruschka 2010, S. 78).

Schulinspektion als Steuerungsinstrument ist ebenso hinterfragbar. So zeigen die angeführten Inter-Beurteiler-Studien ein teilweise beachtliches Maß an Beurteilervarianz. Gerade wenn die Expertenurteile von Inspektorinnen und Inspektoren als problematisch und unzuverlässig gerahmt werden, gibt es zwei gegensätzliche Lösungsansätze. Einerseits kann man versuchen, das wahrgenommene Problem intern zu lösen. Hier böten sich z. B. weitere Beurteilerschulungen und Vereindeutigung der Inspektionsinstrumente an, um eine größere Urteilshomogenität zu erzielen. Diese Entwicklungsstrategie zielt darauf ab, soweit wie möglich einheitliche Bewertungen sicherzustellen. Hierdurch würde primär der Kontrollfunktion von Schulinspektion gedient.

Andererseits kann Schulinspektion aufgrund vermeintlicher Defizite für nicht reparabel erklärt werden. Die Kontrollfunktion kann auch über andere Monitoring-Verfahren realisiert werden. Schulinspektion könnte z. B. durch standardisierte Leistungstests ersetzt werden, da letztere über eine vermeintlich höhere wissenschaftliche Güte verfügen. So führt Schulinspektion in den USA bspw. ein Schattendasein im Vergleich zur dort dominanten psychometrischen Testbranche.

Eine andere Rahmung wird in den referierten qualitativen Studien vorgenommen. Expertenurteile werden, da die Bedeutung von Urteilshomogenität relativiert wird, nicht als Schwäche, sondern als eine potenzielle Stärke von Schulinspektionsverfahren gesehen. Gerade Expertenurteile vermögen so, komplexe Phänomene wie Schul- und Unterrichtsqualität besser zu fassen, als dies etwa mittels Leistungstests für Schülerinnen und Schüler möglich wäre. Inspektorinnen und Inspektoren können bei ihren Bewertungen auf ihr implizites Wissen zurückgreifen, das sie im Inspektionsalltag, aber ggf. auch als Lehrkraft und in Leitungsfunktionen erworben haben (vgl. Bromme 2004; Polanyi 1985).

Auf der „Hinterbühne“ darf die vollständige Expertise zur Geltung kommen, auch wenn sie nicht bis ins letzte Detail den expliziten Verfahrensschritten und Vorschriften entsprechen sollte. Eine dies unterstützende Entwicklungsstrategie

würde beinhalten, den fachlichen Austausch zwischen den Inspektorinnen und Inspektoren zu fördern. Ziel wäre es, eine *Community of Practice* bzw. professionelle Lerngemeinschaft entstehen zu lassen, in der Wissen geteilt sowie Probleme thematisiert und bearbeitet werden.

Auf der „Vorderbühne“ wird der Form halber weiterhin ein abgesichertes Ergebnis in einem standardisierten Format präsentiert werden müssen. Wie valide ein solches Ergebnis ist, ist dabei möglicherweise sekundär. Denn auch das valideste Ergebnis bleibt für Schul- und Unterrichtsentwicklung womöglich folgenlos. Wichtiger scheint, zumindest wenn die Entwicklungsfunktion von Schulinspektion im Vordergrund steht, ob es den Inspektorinnen und Inspektoren gelingt, den schulischen Akteuren zu neuen Einsichten zu verhelfen, für Veränderungsbedarf zu sensibilisieren und für Entwicklungsaufgaben zu motivieren.

Literatur

- Altrichter, H., Ehren, M. C. M., McNamara, G. & O'Hara, J. (2014). Wie will Schulinspektion wirken? Analyse von Annahmen über Wirkungsmechanismen von Schulinspektion in sechs europäischen Ländern. In M. Pfeifer (Hrsg.), *Schulqualität und Schulentwicklung. Theorien, Analysen und Potenziale* (S. 184–207). Münster: Waxmann.
- Biesta, G. (2010). Valuing what we measure or measuring what we value? On the need to engage with the question of purpose in educational evaluation, assessment and measurement. In W. Böttcher, J. N. Dicke & N. Hoglebe (Hrsg.), *Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit* (S. 35–46). Münster: Waxmann.
- Borman, G. D., Hewes, G. M., Overman, L. T. & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73 (2), 125–230. Verfügbar unter: <http://doi.org/10.3102/00346543073002125> [Zugriff am 30.03.2014]
- Bos, W., Holtappels, H. G. & Rösner, E. (2006). Schulinspektion in den deutschen Bundesländern – eine Baustellenbeschreibung. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung* (Band 14, S. 81–123). Weinheim: Juventa.
- Böttcher, W. & Keune, M. (2012). Externe Evaluation und die Steuerung der Einzelschule: Kontrolle oder Entwicklung? In M. Ratermann & S. Stöbe-Blossey (Hrsg.), *Governance von Schul- und Elementarbildung* (S. 63–80). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Boyne, G., Day, P. & Walker, R. (2002). The evaluation of public service inspection: A theoretical framework. *Urban Studies*, 39 (7), 1197–1212. Verfügbar unter: <http://doi.org/10.1080/00420980220135563> [Zugriff am 30.03.2014]

- Bromme, R. (2004). Das implizite Wissen des Experten. In B. Koch-Priewe, F.-U. Kolbe & J. Wildt (Hrsg.), *Grundlagenforschung und mikrodidaktische Reformansätze zur Lehrerbildung* (S. 22–48). Bad Heilbrunn: Klinkhardt.
- Cale, L. & Harris, J. (2003). OFSTED revisited: Re-living the experience of an initial teacher training PE inspection. *European Physical Education Review*, 9 (2), 135–161. <http://doi.org/10.1177/1356336X03009002002>
- Campbell, J. & Husbands, C. (2000). On the reliability of OFSTED inspection of initial teacher training: A case study. *British Educational Research Journal*, 26 (1), 39–48. Verfügbar unter: <http://doi.org/10.1080/014119200109507> [Zugriff am 30.03.2014]
- Dederling, K. (2012). Schulinspektion als wirksamer Weg der Systemsteuerung? *Zeitschrift für Pädagogik*, 58 (1), 69–88.
- Döbert, H., Rürup, M. & Dederling, K. (2008). Externe Evaluation von Schulen in Deutschland – die Konzepte der Bundesländer, ihre Gemeinsamkeiten und Unterschiede. In H. Döbert & K. Dederling (Hrsg.), *Externe Evaluation von Schulen. Historische, rechtliche und vergleichende Aspekte* (S. 63–151). Münster: Waxmann.
- Ehren, M. C. M. & Visscher, A. J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54 (1), 51–72. Verfügbar unter: <http://doi.org/10.1111/j.1467-8527.2006.00333.x> [Zugriff am 30.03.2014]
- Eisner, E. W. (2013). The roots of connoisseurship and criticism: A personal journey. In M. C. Alkin (Hrsg.), *Evaluation roots: A wider perspective of theorists' views and influences* (2. Aufl., S. e-1–e-7). Los Angeles: Sage. Verfügbar unter: http://www.sagepub.com/upm-data/47742_alkin2e_ch31and32.pdf [Zugriff am 30.03.2014]
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59 (1), 255–278. Verfügbar unter: <http://doi.org/10.1146/annurev.psych.59.103006.093629> [Zugriff am 30.03.2014]
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fitz-Gibbon, C. T. & Stephenson-Forster, N. J. (1999). Is Ofsted helpful? An evaluation using social science criteria. In C. Cullingford (Hrsg.), *An inspector calls: Ofsted and its effect on school standards* (S. 97–118). London: Kogan Page.
- Gärtner, H. (2013). Praxis und Verhältnis interner und externer Evaluation im Schulsystem im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 16 (4), 693–712. Verfügbar unter: <http://doi.org/10.1007/s11618-013-0444-7> [Zugriff am 30.03.2014]
- Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62 (1), 451–482. Verfügbar unter: <http://doi.org/10.1146/annurev-psych-120709-145346> [Zugriff am 30.03.2014]
- Gigerenzer, G. & Selten, R. (2002). Rethinking rationality. In G. Gigerenzer & R. Selten (Hrsg.), *Bounded rationality: The adaptive toolbox* (S. 1–12). Cambridge, MA: MIT Press.
- Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday Anchor.

- Gray, J. & Wilcox, B. (1995). The methodologies of school inspection: Issues and dilemmas. In T. Brighouse & B. Moon (Hrsg.), *School Inspection* (S. 127–142). London: Pitman.
- Gruschka, A. (2010). Die Schulinspektion war da und hinterließ einen Bericht. *Pädagogische Korrespondenz*, 23 (41), 75–92.
- Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York, NY: Oxford University Press.
- Heid, H. (2013). Logik, Struktur und Prozess der Qualitätsbeurteilung von Schule und Unterricht. *Zeitschrift für Erziehungswissenschaft*, 16 (2), 405–431. Verfügbar unter: <http://doi.org/10.1007/s11618-013-0363-7> [Zugriff am 30.03.2014]
- Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation. *Zeitschrift für Erziehungswissenschaft*, 14 (2), 259–282. Verfügbar unter: <http://doi.org/10.1007/s11618-011-0204-5> [Zugriff am 30.03.2014]
- IfBQ (Hrsg.) (2011). *Jahresbericht der Schulinspektion Hamburg. 2009–2010*. Hamburg: Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ). Verfügbar unter: <http://epub.sub.uni-hamburg.de/epub/volltexte/2011/10666/> [Zugriff am 30.03.2014]
- IfBQ (Hrsg.) (2012). *Jahresbericht der Schulinspektion, Schuljahr 2010/2011*. Hamburg: Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ). Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:101:1-2013101511516> [Zugriff am 30.03.2014]
- IfBQ (2014). FAQ Schulinspektion. Verfügbar unter: <https://www.hamburg.de/bsb/faq/> [Zugriff am 30.03.2014]
- Jagd, S. (2011). Pragmatic sociology and competing orders of worth in organizations. *European Journal of Social Theory*, 14 (3), 343–359. Verfügbar unter: <http://doi.org/10.1177/1368431011412349> [Zugriff am 30.03.2014]
- Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50 (3), 456–460. Verfügbar unter: <http://doi.org/10.1518/001872008X288385> [Zugriff am 30.03.2014]
- Klein, G., Calderwood, R. & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, 4 (3), 186–209. Verfügbar unter: <http://doi.org/10.1518/155534310X12844000801203> [Zugriff am 30.03.2014]
- Kotthoff, H.-G. & Böttcher, W. (2010). Neue Formen der „Schulinspektion“: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 15–39). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kundin, D. M. (2010). A conceptual framework for how evaluators make everyday practice decisions. *American Journal of Evaluation*, 31 (3), 347–362. Verfügbar unter: <http://doi.org/10.1177/1098214010366048> [Zugriff am 30.03.2014]
- Lambrecht, M. & Rürup, M. (2012). Bildungsforschung im Rahmen einer evidence based policy: Das Beispiel „Schulinspektion“. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung* (S. 57–77). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38 (1), 201–221. Verfügbar unter: <http://doi.org/10.1146/annurev-soc-070308-120022> [Zugriff am 30.03.2014]
- Lincoln, Y.S., Lynham, S.A. & Guba, E.G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N.K. Denzin & Y.S. Lincoln (Hrsg.), *The Sage handbook of qualitative research* (4. Aufl., S. 97–128). Thousand Oaks, CA: Sage.
- Lindgren, J. (2015). The front and back stages of Swedish school inspection: Opening the black box of judgment. *Scandinavian Journal of Educational Research*, 59 (1), 58–76. Verfügbar unter: <http://doi.org/10.1080/00313831.2013.838803> [Zugriff am 30.03.2014]
- Matthews, P., Holmes, J.R., Vickers, P. & Corporaal, B. (1998). Aspects of the reliability and validity of school inspection judgements of teaching quality. *Educational Research and Evaluation*, 4 (2), 167–188. Verfügbar unter: <http://doi.org/10.1076/edre.4.2.167.6959> [Zugriff am 30.03.2014]
- Meyer, H. (2012). Theoretische Grundlagen der Inspektionstätigkeit. In I. Oldenburg (Hrsg.), *Schule und Inspektion. 9 kritische Studien* (S. 17–64). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Millett, A. & Johnson, D.C. (1998). Expertise or ‘baggage’? What helps inspectors to inspect primary mathematics? *British Educational Research Journal*, 24 (5), 503–518. Verfügbar unter: <http://doi.org/10.1080/0141192980240502> [Zugriff am 30.03.2014]
- Müller, S. & Pietsch, M. (2011). Was wir messen, wenn wir Unterrichtsqualität messen. Inter-Beurteilerübereinstimmungen und -Reliabilität bei Unterrichtsbeobachtungen im Rahmen von Schulinspektion. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht* (S. 33–55). Münster: Waxmann.
- Nixon, J. & Rudduck, J. (1993). The role of professional judgment in the local inspection of schools: A study of six local education authorities. *Research Papers in Education*, 8 (2), 135–148. Verfügbar unter: <http://doi.org/10.1080/0267152930080202> [Zugriff am 30.03.2014]
- Nutley, S., Levitt, R., Solesbury, W. & Martin, S. (2012). Scrutinizing performance: How assessors reach judgements about public services. *Public Administration*, 90 (4), 869–885. Verfügbar unter: <http://doi.org/10.1111/j.1467-9299.2011.02022.x> [Zugriff am 30.03.2014]
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 13 (1), 121–148.
- Pietsch, M. & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 11 (3), 430–452. Verfügbar unter: <http://doi.org/10.1007/s11618-008-0021-7> [Zugriff am 30.03.2014]
- Polanyi, M. (1985). *Implizites Wissen*. Frankfurt a. M.: Suhrkamp.
- Power, M. (1999). *The audit society: Rituals of verification*. Oxford: Oxford University Press.

- Praetorius, A.-K. (2013). Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter – Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31 (2), 174–185.
- Preuß, B., Wissinger, J. & Brüsemeister, T. (2015). Einführung der Schulinspektion: Struktur und Wandel regionaler Governance im Schulsystem. In H.J. Abs, T. Brüsemeister, M. Schemmann & J. Wissinger (Hrsg.), *Governance im Bildungssystem* (S. 117–142). Wiesbaden: Springer VS Verlag für Sozialwissenschaften.
- Prior, L. (2003). *Using documents in social research*. London: Sage.
- Rürup, M. (2008). Typen der Schulinspektion in den deutschen Bundesländern. *Die Deutsche Schule*, 100 (4), 467–477.
- Sadler, D.R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34 (2), 159–179. Verfügbar unter: <http://doi.org/10.1080/02602930801956059> [Zugriff am 30.03.2014]
- Sadler, D.R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, 67 (3), 273–288. Verfügbar unter: <http://doi.org/10.1007/s10734-013-9649-1> [Zugriff am 30.03.2014]
- Scheerens, J., Glas, C. & Thomas, S. (2003). *Educational evaluation, assessment, and monitoring: A systemic approach*. Lisse: Swets & Zeitlinger.
- Simon, H.A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, 1 (1), 25–39. Verfügbar unter: <http://doi.org/10.1007/BF02512227> [Zugriff am 30.03.2014]
- Sinkinson, A. & Jones, K. (2001). The validity and reliability of OFSTED: Judgements of the quality of secondary mathematics initial teacher education courses. *Cambridge Journal of Education*, 31 (2), 221. Verfügbar unter: <http://doi.org/10.1080/0305764012006155> [Zugriff am 30.03.2014]
- Sommer, N. (2011). Wie beurteilen schulische Gruppen die erlebte Schulinspektion? Ergebnisse einer Befragung. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht* (S. 137–164). Münster: Waxmann.
- Sowada, M.G. & Dederich, K. (2014). Ermessensspielräume in der Bewertungsarbeit von Schulinspektor/innen. *Zeitschrift für Bildungsforschung*, 4 (2), 119–135. Verfügbar unter: <http://doi.org/10.1007/s35834-014-0091-y> [Zugriff am 30.03.2014]
- Strathern, M. (Hrsg.) (2000). *Audit cultures: Anthropological studies in accountability, ethics and the academy*. London: Routledge.
- Tourmen, C. (2009). Evaluators' decision making: The relationship between theory, practice, and experience. *American Journal of Evaluation*, 30 (1), 7–30. Verfügbar unter: <http://doi.org/10.1177/1098214008327602> [Zugriff am 30.03.2014]
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185 (4157), 1124–1131. Verfügbar unter: <http://doi.org/10.1126/science.185.4157.1124> [Zugriff am 30.03.2014]

- Wilcox, B. & Gray, J. (1996). The methodological problems of inspection. *Inspecting schools: Holding schools to account and helping schools to improve* (S. 65–80). Buckingham: Open University Press.
- Wilson, T.A. (1996). *Reaching for a better standard: English school inspection and the dilemma of accountability for American public schools*. New York, NY: Teachers College Press.